

# DISTRIBUTIONAL CONVERGENCE FOR THE NUMBER OF SYMBOL COMPARISONS USED BY QUICKSELECT

JAMES ALLEN FILL AND TAKEHIKO NAKAMA

## ABSTRACT

When the search algorithm **QuickSelect** compares keys during its execution in order to find a key of target rank, it must operate on the keys' representations or internal structures, which were ignored by the previous studies that quantified the execution cost for the algorithm in terms of the number of required key comparisons. In this paper, we analyze running costs for the algorithm that take into account not only the number of key comparisons but also the cost of each key comparison. We suppose that keys are represented as sequences of symbols generated by various probabilistic sources and that **QuickSelect** operates on individual symbols in order to find the target key. We identify limiting distributions for the costs and derive integral and series expressions for the expectations of the limiting distributions. These expressions are used to recapture previously obtained results on the number of key comparisons required by the algorithm.

## 1. INTRODUCTION AND SUMMARY

**QuickSelect**, introduced by Hoare [11] in 1961 and also known as **Find** or “Hoare’s selection algorithm”, is a simple search algorithm widely used for finding a key (an object drawn from a linearly ordered set) of target rank in a file of keys. We briefly review the operation of the algorithm. Suppose that there are  $n$  keys (we will suppose that these are all distinct) and that the target rank is  $m$ , where  $1 \leq m \leq n$ . **QuickSelect**  $\equiv$  **QuickSelect**( $n, m$ ) chooses a uniformly random key, called the pivot, and compares each other key to it. This determines the rank  $j$  (say) of the pivot. If  $j = m$ , then the algorithm returns the pivot key and terminates. If  $j > m$ , then **QuickSelect** is applied recursively to find the key of rank  $m$  in the set of  $j - 1$  keys found to be smaller than the pivot. If  $j < m$ , then **QuickSelect** is applied recursively to find the key of rank  $m - j$  in the set of  $n - j$  keys larger than the pivot.

Many studies have examined this algorithm to quantify its execution costs (a non-exhaustive list of references is Knuth [13]; Mahmoud, Modarres, and Smythe [15]; Prodinger [18]; Grübel and Rösler [10]; Lent and Mahmoud [14]; Grübel [9]; Mahmoud and Smythe [16]; Devroye [5]; Hwang and Tsai [12]; Fill and Nakama [8]; and Vallée, Clément, Fill, and Flajolet [23]); and all of them except for Fill and Nakama [8] and Vallée *et al.* [23] have conducted the quantification with regard to the number of key comparisons required by the algorithm to achieve its task. As a

---

*Date:* February 10, 2012.

Research for both authors supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.

result, most of the theoretical results on the complexity of **QuickSelect** are about expectations or distributions for the number of required key comparisons.

However, one can reasonably argue that analyses of **QuickSelect** in terms of the number of key comparisons cannot fully quantify its complexity. For instance, if keys are represented as binary strings, then individual bits of the strings must be compared in order for **QuickSelect** to complete its task, and results obtained by analyzing the algorithm with respect to the number of bit comparisons required to find a target key more accurately reflect actual execution costs. (We will consider bit comparisons as an example of symbol comparisons.) When **QuickSelect** (or any other algorithm) compares keys during its execution, it must operate on the keys' representations or internal structures, so these should not be ignored in fully characterizing the performance of the algorithm. Also, symbol-complexity analysis allows us to compare key-based algorithms such as **QuickSelect** and **QuickSort** with digital algorithms such as those utilizing digital search trees.

Fill and Janson [7] pioneered symbol-complexity analysis by analyzing the expected number of bit comparisons required by **QuickSort**. They assumed that the algorithm is applied to keys that are i.i.d. (independent and identically distributed) from the uniform distribution over  $(0, 1)$  and represented (via their binary expansions) as binary strings, and that the algorithm operates on individual bits in order to do comparisons and find the target key. They found that the expected number of bit comparisons required by **QuickSort** to sort  $n$  keys is asymptotically equivalent to  $n(\ln n)(\lg n)$  (where  $\lg$  denotes binary logarithm), whereas the lead-order term of the expected number of *key* comparisons is  $2n \ln n$ , smaller by a factor of order  $\log n$ . In their Section 6 they also considered i.i.d. keys drawn from other distributions with density on  $(0, 1)$ .

By closely following [7], Fill and Nakama [8] studied the expected number of bit comparisons required by **QuickSelect**. More precisely, they treated the case of i.i.d. uniform keys represented as binary strings and produced exact expressions for the expected number of bit comparisons by **QuickSelect** $(n, m)$  for general  $n$  and  $m$ . Their asymptotic results were limited to the algorithms **QuickMin**, **QuickMax**, and **QuickRand**. Here **QuickMin** refers to **QuickSelect** applied to find the smallest key, i.e., to **QuickSelect** $(n, m)$  with  $m = 1$ ; and **QuickMax** similarly refers to **QuickSelect** $(n, m)$  with  $m = n$ . **QuickRand** is the algorithm that results from taking  $m$  to be uniformly distributed over  $\{1, 2, \dots, n\}$ . They showed that the expected number of bit comparisons required by **QuickMin** or **QuickMax** is asymptotically linear in  $n$  with lead-order coefficient approximately equal to 5.27938. Thus in these cases the expected number of bit comparisons is asymptotically larger than that of *key* comparisons required to complete the same task only by a constant factor, since the expectation for key comparisons is asymptotically  $2n$ . Fill and Nakama [8] also found that the expected number of bit comparisons required by **QuickRand** is also asymptotically linear in  $n$  (with slope approximately 8.20731), as for key comparisons (with slope 3).

Vallée *et al.* [23] extended the average-case analyses of [7] and [8] to keys represented by sequences of general symbols generated by any of a wide variety of sources that include memoryless, Markov, and other dynamical sources. They broadly extended the results of [8] in another direction as well by treating **QuickQuant** $(n, \alpha)$  for general  $\alpha \in [0, 1]$ , not just **QuickMin**, **QuickMax**, and **QuickRand**. Here the algorithm **QuickQuant** $(n, \alpha)$  (for “Quick Quantile”) refers to **QuickSelect** $(n, m_n)$  with

$m_n/n \rightarrow \alpha$ . Roughly summarized, Vallée *et al.* showed that if symbols are generated by a suitably nice source, then the expected number of symbol comparisons in processing a file of  $n$  keys is of order  $n \log^2 n$  for **QuickSort** and, for any  $\alpha$ , of order  $n$  for **QuickQuant**( $n, \alpha$ ). (For example, all memoryless sources are suitably nice.) For a more detailed discussion of sources and the results of Vallée *et al.* [23] for **QuickQuant**, see Section 2.

The main purpose of this paper is to extend the average-case analysis of Vallée *et al.* [23] by establishing limiting distributions for the number of symbol comparisons. To our knowledge the present paper is the first to establish a limiting distribution for the number of symbol comparisons required by any key-based algorithm. Our elementary approach allows us to handle rather general kinds of “cost” for comparing two keys, and in particular to recover in a rather direct way known results about the number of key comparisons. There is no disadvantage to allowing general costs, since our results rely on at most broad limitations on the nature of the cost.

**Outline of the paper.** We shall be concerned primarily with **QuickQuant**  $\equiv$  **QuickQuant**( $n, \alpha$ ), which is what we call the algorithm **QuickSelect** when applied to find the key of rank  $m_n$  in a file of size  $n$ , where we are given  $0 \leq \alpha \leq 1$  and a sequence  $(m_n)$  such that  $m_n/n \rightarrow \alpha$ . It turns out to be convenient mathematically to analyze a close cousin to **QuickQuant** introduced by Vallée *et al.* [23], namely, **QuickVal**, and then treat **QuickQuant** by comparison. So, after a careful description of the probabilistic models used to govern the generation of keys in Section 2.1, a review of known results about key and symbol comparisons in Section 2.2, and a description of **QuickVal** in Section 2.3, in Section 3 we establish limiting-distribution results for **QuickVal** (whose main theorems are Theorem 3.1 and Theorem 3.4) and then move on to **QuickQuant** in Section 4 (which contains Theorem 5.1, the main theorem of this paper).

Subsequent to the research leading to the present paper, and using a rather different approach, the first author [6] has found a limiting distribution for the number of symbol comparisons used by **QuickSort** for a wide variety of probabilistic sources.

**Remark 1.1.** Although the contraction method has been used in finding limiting distributions for the number of key comparisons required by recursive algorithms such as **QuickSort** (e.g., Rösler [20], Rösler and Rüschendorf [21]), our analysis does not depend on it. In examining convergence for the number of key comparisons used by **QuickQuant**, Grübel and Rösler [10] mentioned that they did not use the contraction method due to the parameter that represents target rank. (However, they did engage in contraction arguments to characterize the limiting distribution.) Interestingly, Mahmoud *et al.* [15] succeeded in establishing a fixed point equation to identify the limiting distributions of the normalized numbers of key comparisons required by **QuickRand**, **QuickMin**, and **QuickMax**. Régnier [19] used martingales to show convergence for the number of key comparisons required by **QuickSort**.

## 2. BACKGROUND AND PRELIMINARIES

**2.1. Probabilistic source models for the keys.** In this subsection we describe what is meant by a probabilistic source, our model for how the i.i.d. keys are generated, using the terminology and notation of Vallée *et al.* [23].

Let  $\Sigma$  denote a totally ordered alphabet (set of symbols), assumed to be isomorphic either to  $\{0, \dots, r-1\}$  for some finite  $r$  or to the full set of nonnegative integers, in either case with the natural order; a *word* is then an element of  $\Sigma^\infty$ , i.e., an infinite sequence (or “string”) of symbols. We will follow the customary practice of denoting a word  $w = (w_1, w_2, \dots)$  more simply by  $w_1 w_2 \dots$ .

We will use the word “prefix” in two closely related ways. First, the symbol strings belonging to  $\Sigma^k$  are called *prefixes* of length  $k$ , and so  $\Sigma^* := \bigcup_{0 \leq k < \infty} \Sigma^k$  denotes the set of all prefixes of any nonnegative finite length. Second, if  $w = w_1 w_2 \dots$  is a word, then we will call

$$(2.1) \quad w(k) := w_1 w_2 \dots w_k \in \Sigma^k$$

its *prefix of length  $k$* .

*Lexicographic order* is the linear order (to be denoted in the strict sense by  $\prec$  and in the weak sense by  $\preceq$ ) on the set of words specified by declaring that  $w \prec w'$  if (and only if) for some  $0 \leq k < \infty$  the prefixes of  $w$  and  $w'$  of length  $k$  are equal but  $w_{k+1} < w'_{k+1}$ . We denote the *cost* of determining  $w \prec w'$  when comparing distinct words  $w$  and  $w'$  by  $c(w, w')$ ; we will always assume that the function  $c$  is symmetric and nonnegative.

*Example 2.1.* Here is an example of a natural class of cost functions. Start with nonnegative symmetric functions  $c_i : \Sigma \times \Sigma \rightarrow [0, \infty)$ ,  $i = 1, 2, \dots$ , modeling the cost of comparing symbols in the respective  $i$ th positions of two words. This allows for the symbol-comparison costs to depend both on the positions of the symbols in the words and on the symbols themselves. Then, for comparisons of distinct words, define

$$(2.2) \quad c(w, w') := \sum_{i=1}^{k+1} c_i(w_i, w'_i) = \sum_{i=1}^k c_i(w_i, w_i) + c_{k+1}(w_{k+1}, w'_{k+1})$$

where  $k$  is the length of the longest common prefix of  $w$  and  $w'$ .

(a) If  $c_i \equiv \delta_{i_0, i}$  (independent of the symbols being compared) for given positive integer  $i_0$ , then  $c$  is the cost used in counting comparisons of symbols in position  $i_0$ . (For example, if  $i_0 = 1$  then  $c \equiv 1$  is the cost used in counting key comparisons.) Observe that all finite linear combinations of such cost functions  $\delta_{i_0, \cdot}$  are of the form (2.2), and therefore, by the Cramér–Wold device (e.g., [1, Section 29]), if  $S_{i_0}$  denotes the total number of comparisons of symbols in position  $i_0$ , then the *joint* distribution of  $(S_1, S_2, \dots)$  can (at least in principle) be obtained by studying cost functions of the form (2.2).

(b) If  $c_i \equiv 1$  for all  $i$ , then  $c \equiv k + 1$  is the cost used in counting symbol comparisons.

A *probabilistic source* is simply a stochastic process  $W = W_1 W_2 \dots$  with state space  $\Sigma$  (endowed with its total  $\sigma$ -field) or, equivalently, a random variable  $W$  taking values in  $\Sigma^\infty$  (with the product  $\sigma$ -field). According to Kolmogorov’s consistency criterion (e.g., [2, Theorem 3.3.6]), the distributions  $\mu$  of such processes are in one-to-one correspondence with consistent specifications of finite-dimensional marginals, that is, of the probabilities

$$p_w := \mu(\{w_1 \dots w_k\} \times \Sigma^\infty), \quad w = w_1 w_2 \dots w_k \in \Sigma^*.$$

Here the *fundamental probability*  $p_w$  is the probability that a word drawn from  $\mu$  has  $w_1 \dots w_k$  as its length- $k$  prefix.

Because the analysis of **QuickSelect** is significantly more complicated when its input keys are not all distinct, we will restrict attention to probabilistic sources with continuous distributions  $\mu$ . Expressed equivalently in terms of fundamental probabilities, our continuity assumption is that for any  $w = w_1 w_2 \cdots \in \Sigma^\infty$  we have  $p_{w(k)} \rightarrow 0$  as  $k \rightarrow \infty$ , recalling the prefix notation (2.1).

*Example 2.2.* We present a few classical examples of sources. For more examples, and for further discussion, see Section 3 of [23].

(a) In computer science jargon, a *memoryless source* is one with  $W_1, W_2, \dots$  i.i.d. Then the fundamental probabilities  $p_w$  have the product form

$$p_w = p_{w_1} p_{w_2} \cdots p_{w_k}, \quad w = w_1 w_2 \cdots w_k \in \Sigma^*.$$

(b) A *Markov source* is one for which  $W_1 W_2 \cdots$  is a Markov chain.

(c) An intermittent source over the finite alphabet  $\Sigma = \{0, \dots, r-1\}$  models long-range dependence of the symbols within a key and is defined by specifying the conditional distributions  $\mathcal{L}(W_j | W_1, \dots, W_{j-1})$  in a way that pays special attention to a particular symbol  $\underline{\sigma}$ . The source is said to be *intermittent of exponent  $\gamma > 0$  with respect to  $\underline{\sigma}$*  if  $\mathcal{L}(W_j | W_1, \dots, W_{j-1})$  depends only on the maximum value  $k$  such that the last  $k$  symbols in the prefix  $W_1 \cdots W_{j-1}$  are all  $\underline{\sigma}$  and (i) is the uniform distribution on  $\Sigma$ , if  $k = 0$ ; and (ii) if  $1 \leq k \leq j-1$ , assigns mass  $[k/(k+1)]^\gamma$  to  $\underline{\sigma}$  and distributes the remaining mass uniformly over the remaining elements of  $\Sigma$ .

We next present an equivalent description of probabilistic sources (with a corresponding equivalent condition for continuity) that will prove convenient because it allows us to treat all sources within a uniform framework. If  $M$  is any measurable mapping from  $(0, 1)$  (with its Borel  $\sigma$ -field) into  $\Sigma^\infty$  and  $U$  is distributed  $\text{unif}(0, 1)$ , then  $M(U)$  is a probabilistic source. Conversely, given any probability measure  $\mu$  on  $\Sigma^\infty$  there exists a monotone measurable mapping  $M$  such that  $M(U)$  has distribution  $\mu$  when  $U \sim \text{unif}(0, 1)$ ; here (weakly) *monotone* means that  $M(t) \preceq M(u)$  whenever  $t \leq u$ . Indeed, if  $F$  is the distribution function

$$F(w) := \mu\{w' \in \Sigma^\infty : w' \preceq w\}, \quad w \in \Sigma^\infty,$$

for  $\mu$ , then we can always use the inverse probability transform

$$M(u) := \inf\{w \in \Sigma^\infty : u \preceq F(w)\}, \quad u \in (0, 1)$$

for  $M$ . The measure  $\mu$  is continuous if and only if this  $M$  is strictly monotone.

So henceforth we will assume that our keys are generated as  $M(U_1), \dots, M(U_n)$ , where  $M : (0, 1) \rightarrow \Sigma^\infty$  is strictly monotone and  $U_1, \dots, U_n$  (we will call these the “seeds” of the keys) are i.i.d.  $\text{unif}(0, 1)$ . Given a specification of costs  $c(w, w')$  in comparing words, we can now define a source-specific notion of cost by setting

$$\beta(u, t) := c(M(u), M(t)).$$

In our main application,  $\beta_{\text{symb}}(u, t)$  represents the number of symbol comparisons required to compare words with seeds  $u$  and  $t$ .

The following associated terminology and notation from [23] will also prove useful. For each prefix  $w \in \Sigma^*$ , we let  $\mathcal{I}_w = (a_w, b_w)$  denote the interval that contains all seeds whose corresponding words begin with  $w$  and  $\mu_w := (a_w + b_w)/2$  its mid-point. We call  $\mathcal{I}_w$  the *fundamental interval* associated with  $w$ . (There is no need to be fussy as to whether the interval is open or closed or half-open, because the probability that a random seed  $U$  takes any particular value is 0. Also, we always

assume that  $a_w < b_w$ , since the case that  $a_w = b_w$  will not concern us.) The fundamental probability  $p_w$  can be expressed as  $b_w - a_w$ . The *fundamental triangle* of prefix  $w$ , denoted by  $\mathcal{T}_w$ , is the triangular region

$$\mathcal{T}_w := \{(u, t) : a_w < u < t < b_w\},$$

and when  $w$  is the empty prefix we denote this triangle by  $\mathcal{T}$ :

$$\mathcal{T} := \{(u, t) : 0 < u < t < 1\}.$$

For some of our results, the quantity

$$(2.3) \quad \pi_k := \max\{p_w : w \in \Sigma^k\}$$

will play an important role. The following definition of a  $\Pi$ -tamed probabilistic source is taken (with slight modification) from [23]:

**Definition 2.3.** Let  $0 < \gamma < \infty$  and  $0 < A < \infty$ . We say that the source is  $\Pi$ -tamed (with parameters  $\gamma$  and  $A$ ) if the sequence  $(\pi_k)$  at (2.3) satisfies

$$\pi_k \leq A(k+1)^{-\gamma} \text{ for every } k \geq 0.$$

Observe that a  $\Pi$ -tamed source is always continuous. There is a related condition for cost functions  $\beta$  that will be assumed (for suitable values of the parameters) in some of our results:

**Definition 2.4.** Let  $0 < \epsilon < \infty$  and  $0 < c < \infty$ . We say that the symmetric cost function  $\beta \geq 0$  is tamed (with parameters  $\epsilon$  and  $c$ ) if

$$\beta(u, t) \leq c(t - u)^{-\epsilon} \text{ for all } (u, t) \in T.$$

We say that  $\beta$  is  $\epsilon$ -tamed if it is tamed with parameters  $\epsilon$  and  $c$  for some  $c$ .

We leave it to the reader to make the simple verification that a source is  $\Pi$ -tamed with parameters  $\gamma$  and  $A$  if and only if  $\beta_{\text{symb}}$  is tamed with parameters  $\epsilon = 1/\gamma$  and  $c = A^{1/\gamma}$ .

**Remark 2.5.** (a) Many common sources have geometric decrease in  $\pi_k$  (call these “g-tamed”) and so for *any*  $\gamma$  are  $\Pi$ -tamed with parameters  $\gamma$  and  $A$  for suitably chosen  $A \equiv A_\gamma$  [equivalently, the symbol-comparisons cost  $\beta_{\text{symb}}$  is  $\epsilon$ -tamed for any  $\epsilon$ ; in fact, if  $\pi_k \leq b^{-k}$  for every  $k$ , then

$$\beta_{\text{symb}}(u, t) \leq 1 + \log_b \frac{1}{t - u} \text{ for all } (u, t) \in \mathcal{T}].$$

For example, a memoryless source satisfies  $\pi_k = p_{\max}^k$ , where

$$p_{\max} := \sup_{w \in \Sigma^1} p_w$$

satisfies  $p_{\max} < 1$  except in the highly degenerate case of an essentially single-symbol alphabet. We also have  $\pi_k \leq p_{\max}^k$  for any Markov source, where now  $p_{\max}$  is the supremum of all one-step transition probabilities, and so such a source is g-tamed provided  $p_{\max} < 1$ . Expanding dynamical sources (cf. Clément, Flajolet, and Vallée [3]) are also g-tamed.

(b) For an intermittent source as in Example 2.2, for all large  $k$  the maximum probability  $\pi_k$  is attained by the word  $\underline{\sigma}^k$  and equals

$$\pi_k = r^{-1} k^{-\gamma}.$$

Intermittent sources are therefore examples of  $\Pi$ -tamed sources for which  $\pi_k$  decays at a truly inverse-polynomial rate, not an exponential rate as in the case of  $g$ -tamed sources.

**2.2. Known results for the numbers of key and symbol comparisons.** In this subsection we give for **QuickSelect** an abbreviated review of what is already known about the distribution of the number of key comparisons ( $\beta \equiv 1$  in our notation) and (from Vallée *et al.* [23]) about the expected number of symbol comparisons ( $\beta = \beta_{\text{symb}}$ ). To our knowledge, no other cost functions have previously been considered, nor has there been any treatment of the full distribution of the number of symbol comparisons.

Let  $K_{n,m}$  denote the number of key comparisons required by the algorithm to find a key of rank  $m$  in a file of  $n$  keys (with  $1 \leq m \leq n$ ). Thus  $K_{n,1}$  and  $K_{n,n}$  represent the key comparison costs required by **QuickMin** and **QuickMax**, respectively. (Clearly  $K_{n,1} \stackrel{L}{=} K_{n,n}$ ). It has been shown (see Mahmoud *et al.* [15], Hwang and Tsai [12]) that as  $n \rightarrow \infty$ ,  $K_{n,1}/n$  converges in law to the Dickman distribution, which can be described as the distribution of the perpetuity

$$1 + \sum_{k \geq 1} U_1 \cdots U_k,$$

where  $U_k$  are i.i.d. uniform(0,1). Mahmoud *et al.* [15] established a fixed-point equation for the limiting distribution of the normalized (by dividing by  $n$ ) number of key comparisons required by **QuickRand** and also explicitly identified this limiting distribution.

By using process-convergence techniques, Grübel and Rösler [10, Theorem 8] identified, for each  $0 \leq \alpha < 1$ , a nondegenerate random variable  $K(\alpha)$  to which  $K_{n, \lfloor \alpha n \rfloor + 1}/n$  converges in distribution; see also the fixed-point equation in their Theorem 10, and Grübel [9], who used a Markov chain approach and characterized the limiting distribution in his Theorem 3. Earlier, Devroye [4] had shown that

$$\sup_{n \geq 1} \max_{1 \leq m \leq n} \mathbf{P}(K_{n,m} \geq tn) \leq C\rho^t$$

for any  $\rho > 3/4$  and some  $C \equiv C(\rho)$ .

Concerning moments, Grübel and Rösler [10, Theorem 11] showed that  $\mathbf{E} K(\alpha) = 2[1 - \alpha \ln \alpha - (1 - \alpha) \ln(1 - \alpha)]$  and Paulsen [17] calculated higher-order moments of  $K(\alpha)$ . Grübel [9, end of Section 2] proved convergence of the moments for finite  $n$  to the corresponding moments of the limiting  $K(\alpha)$ .

Prior to the present paper, only expectations have been studied for the number of symbol comparisons for **QuickQuant**. The current state of knowledge is summarized by part (i) of Theorem 2 in Vallée *et al.* [23] (see also their accompanying Figures 1–3); we refer the reader to [23] for the other parts of the theorem, which routinely specialize part (i) to **QuickMin**, **QuickMax**, and **QuickRand**.

To review their result we need the notation and terminology of Section 2.1 and a bit more. Using the non-standard abbreviations  $y^+ := (1/2) + y$  and  $y^- := (1/2) - y$  and the convention  $0 \ln 0 := 0$ , we define

$$H(y) := \begin{cases} -(y^+ \ln y^+ + y^- \ln y^-), & \text{if } 0 \leq y \leq 1/2 \\ y^- (\ln y^+ - \ln |y^-|), & \text{if } y \geq 1/2 \end{cases}$$

and then set  $L(y) := 2[1 + H(y)]$ . According to Theorem 2(i) in [23], for any  $\Pi$ -tamed source the mean number of symbol comparisons for **QuickQuant**( $n, \alpha$ ) is



asymptotically  $\rho n + O(n^{1-\delta})$  for some  $\delta > 0$ . Here  $\rho \equiv \rho(\alpha)$  and  $\delta$  both depend on the probabilistic source, with

$$(2.4) \quad \rho := \sum_{w \in \Sigma^*} p_w L \left( \left| \frac{\alpha - \mu_w}{p_w} \right| \right).$$

They derive (2.4) by first proving the equality

$$(2.5) \quad \rho = \int_{\mathcal{T}} \beta(u, t) [(\alpha \vee t) - (\alpha \wedge u)]^{-1} du dt$$

for  $\Pi$ -tamed sources with  $\gamma > 1$ .

**2.3. QuickQuant and QuickVal.** Let  $S_n^Q \equiv S_n^Q(\alpha)$  denote the total cost required by **QuickQuant**( $n, \alpha$ ). To prove convergence of  $S_n^Q/n$  (in suitable senses to be made precise later), we exploit an idea introduced by Vallée *et al.* [23] and begin with the study of a related algorithm, called **QuickVal**  $\equiv$  **QuickVal**( $n, \alpha$ ), which we now describe. **QuickVal** is admittedly somewhat artificial and inefficient; it is important to keep in mind that we study it mainly as an aid to studying **QuickQuant**.

Having generated  $n$  seeds and then  $n$  keys  $M_1, \dots, M_n$  (say) using our probabilistic source, **QuickVal** is a recursive randomized algorithm to find the rank of the additional word  $M(\alpha)$  in the set  $\{M_1, \dots, M_n, M(\alpha)\}$ ; thus, while **QuickQuant** finds the value of the  $\alpha$ -quantile in the *sample* of keys, **QuickVal** dually finds the rank of the *population*  $\alpha$ -quantile in the augmented set. First, **QuickVal** selects a pivot uniformly at random from the set of keys  $\{M_1, \dots, M_n\}$  and finds the rank of the pivot by (a) comparing the pivot with each of the other keys (we will count these comparisons) and (b) comparing the pivot with  $M(\alpha)$  (we will find it convenient not to count the cost of this comparison in the total cost). With probability one, the pivot key will differ from the word  $M(\alpha)$ . If  $M(\alpha)$  is smaller than the pivot key, then the algorithm operates recursively on the set of keys smaller than the pivot and determines the rank of the word  $M(\alpha)$  in the set  $\mathcal{M}_{\text{smaller}} \cup \{M(\alpha)\}$ , where  $\mathcal{M}_{\text{smaller}}$  denotes the set of keys smaller than the pivot. Similarly, if  $M(\alpha)$  is greater than the pivot key, then the algorithm operates recursively on the set of keys larger than the pivot [together with the word  $M(\alpha)$ ]. Eventually the set of words on which the algorithm operates reduces to the singleton  $\{M(\alpha)\}$ , and the algorithm terminates.

Notice that the operation of **QuickVal** is quite close to that of **QuickQuant**, for the same value of  $\alpha$ ; we expect running costs of the two algorithms to be close, since when  $n$  is large the rank of  $M(\alpha)$  in  $\{M_1, \dots, M_n, M(\alpha)\}$  should be close (in relative error terms) to  $\alpha n$ . In fact, we will show that if  $S_n^V \equiv S_n^V(\alpha)$  denotes the total cost of executing **QuickVal**( $n, \alpha$ ), then  $S_n^Q/n$  and  $S_n^V/n$  have the same limiting distribution, assuming only that the cost function  $\beta$  is  $\epsilon$ -tamed for suitably small  $\epsilon$ . In fact, we will show that when all the random variables  $S_1^Q, S_2^Q, \dots$  and  $S_1^V, S_2^V, \dots$  are strategically defined on a common probability space, then  $S_n^Q/n$  and  $S_n^V/n$  both converge in  $L^p$  to a common limit for  $1 \leq p < \infty$ .

### 3. ANALYSIS OF QuickVal

Following some preliminaries in Section 3.1, in Section 3.2 we show that for  $1 \leq p < \infty$ , a suitably defined  $S_n^V/n$  converges in  $L^p$  to a certain random variable  $S$  (defined at the end of Section 3.1) provided only that  $\mathbf{E} S < \infty$ . We also show that, when the cost function is suitably tamed,  $S_n^Q/n$  converges almost surely to



$S$ ; see Theorem 3.4 in Section 3.3. We derive an integral expression for  $\mathbf{E}S$  valid for a completely general cost function  $\beta$  in Section 3.4 and use it to compute the expectation when  $\beta \equiv 1$ . In Section 3.5, we focus on  $\mathbf{E}S$  with  $\beta = \beta_{\text{sympb}}$  and derive a series expression for the expectation. Few comparisons of results obtained here with the known results reviewed in Section 2.2 are made in the present section; most such comparisons are deferred to (the first paragraph of) Section 4, where the previously-studied algorithm of greater interest, **QuickQuant**, is treated.

**3.1. Preliminaries.** Our goal is to establish a limit, in various senses, for the ratio of the total cost required by **QuickVal** when applied to a file of  $n$  keys to  $n$ . It will be both natural and convenient to define all these total costs, one for each value of  $n$ , in terms of a single infinite sequence  $(U_i)_{i \geq 1}$  of seeds that are i.i.d. uniform(0, 1). Indeed, let  $L_0 := 0$  and  $R_0 := 1$ . For  $k \geq 1$ , inductively define

$$\begin{aligned} (3.1) \quad \tau_k &:= \inf\{i : L_{k-1} < U_i < R_{k-1}\}, \\ (3.2) \quad L_k &:= \mathbf{1}(U_{\tau_k} < \alpha)U_{\tau_k} + \mathbf{1}(U_{\tau_k} > \alpha)L_{k-1}, \\ (3.3) \quad R_k &:= \mathbf{1}(U_{\tau_k} < \alpha)R_{k-1} + \mathbf{1}(U_{\tau_k} > \alpha)U_{\tau_k}, \\ (3.4) \quad S_{n,k} &:= \sum_{i: \tau_k < i \leq n} \mathbf{1}(L_{k-1} < U_i < R_{k-1}) \beta(U_i, U_{\tau_k}). \end{aligned}$$

(Note that  $S_{n,k}$  vanishes if  $\tau_k \geq n$ .) We then claim that, for each  $n$ ,

$$(3.5) \quad S_n^V := \sum_{k \geq 1} S_{n,k}$$

has the distribution of the total cost required by **QuickVal**( $n, \alpha$ ).

We offer some explanation here. For each  $k \geq 1$ , the random interval  $(L_{k-1}, R_{k-1})$  (whose length decreases monotonically in  $k$ ) contains both the target seed  $\alpha$  and the seed  $U_{\tau_k}$  corresponding to the  $k$ th pivot; the interval contains precisely those seed values still under consideration after  $k-1$  pivots have been performed. The only difference between how we have defined  $S_n^V$  and how it is usually defined is that we have chosen the initial pivot seed to be the *first* seed rather than a random one, and have made this same change recursively. But our change is permissible because of the following basic probabilistic fact: If  $U_1, \dots, U_N, M$  are independent random variables with  $U_1, \dots, U_N$  i.i.d. uniform(0, 1) and  $M$  uniformly distributed on  $\{1, \dots, N\}$ , then  $U_M$ , like  $U_1$ , is distributed uniform(0, 1). Thus the conditional distribution of  $U_{\tau_k}$  given  $(L_{k-1}, R_{k-1})$  is uniform( $L_{k-1}, R_{k-1}$ ).

We illustrate our notation for the first two pivots. First,  $\tau_1 = 1$ ; that is, the seed of the first pivot is the uniform(0, 1) random variable  $U_1$ . After that, if  $\alpha < U_1$  then the seed  $U_{\tau_2}$  of the second pivot is chosen as the first seed falling in  $(0, U_1)$ , while if  $\alpha > U_1$  then  $U_{\tau_2}$  is the first seed falling in  $(U_1, 1)$ . We note that if  $\alpha = 0$  (which means that we are dealing with the total cost required by **QuickMin**), then the first of these two cases is always the one that applies and so for every  $k \geq 1$  we have  $L_k = 0$  and  $R_k = U_{\tau_k}$ ; we then have that  $U_{\tau_k}$  is just the  $k$ th record low value among  $U_1, U_2, \dots$ .

In order to describe the limit of  $S_n^V/n$ , we let

$$(3.6) \quad \begin{aligned} I(t, x, y) &:= \int_x^y \beta(u, t) du, \\ I_k &:= I(U_{\tau_k}, L_{k-1}, R_{k-1}), \end{aligned}$$

$$(3.7) \quad S := \sum_{k \geq 1} I_k.$$

Notice that in the case  $\beta \equiv 1$  of key comparisons we have  $I(t, x, y) \equiv y - x$  and so  $I_k = R_{k-1} - L_{k-1}$ .

In Section 3.2 we show for  $1 \leq p < \infty$  that  $S_n^V/n$  converges in  $L^p$  to  $S$  as  $n \rightarrow \infty$  under proper technical conditions. Under a stronger assumption, we will also prove almost sure convergence in Section 3.3.

**3.2. Convergence of  $S_n^V/n$  in  $L^p$  for  $1 \leq p < \infty$ .** Theorem 3.1 is our main result concerning **QuickVal**. To state the result, we need the following notation, extending that of (3.6):

$$(3.8) \quad \begin{aligned} I_p(t, x, y) &:= \int_x^y \beta^p(u, t) du, \\ I_{p,k} &:= I_p(U_{\tau_k}, L_{k-1}, R_{k-1}). \end{aligned}$$

**Theorem 3.1.** *If  $1 \leq p < \infty$  and*

$$(3.9) \quad \sum_{k \geq 1} (\mathbf{E} I_{p,k})^{1/p} < \infty,$$

*then  $S_n^V/n$  converges in  $L^p$  (and therefore also in probability and in distribution) to  $S$  as  $n \rightarrow \infty$ .*

**Remark 3.2.** For  $p = 1$ , notice that the assumption of Theorem 3.1 only requires that  $\mathbf{E} S < \infty$ , which is equivalent to the assertion that  $\sum_{k \geq 1} \mathbf{E} I_k < \infty$ .

*Proof.* We use  $\|\cdot\|$  to denote  $L^p$ -norm. As background, we recall that the  $L^p$  law of large numbers ( $L^p$ LLN) states that for  $1 \leq p < \infty$  and i.i.d. random variables  $\xi_1, \xi_2, \dots$  with finite  $L^p$ -norm, the sample means  $\bar{\xi}_n = n^{-1} \sum_{i=1}^n \xi_i$  converge in  $L^p$  to the expectation. To prove this, we may assume with no loss of generality that the expectation is 0, and then the bound

$$\mathbf{P}(|\bar{\xi}_n|^p > c) \leq c^{-1} \mathbf{E} |\bar{\xi}_n|^p \leq c^{-1} \|\xi_1\|_p^p$$

following from Markov's inequality and the triangle inequality for  $L^p$ -norm shows that the sequence  $(|\bar{\xi}_n|^p)$  is uniformly integrable. So the  $L^p$ LLN follows from the better-known strong law of large numbers.

Returning to the setting of the theorem, fix  $k$ . Conditionally given the quadruple  $C_k = (L_{k-1}, R_{k-1}, \tau_k, U_{\tau_k})$ , the random variables  $U_i$  with  $i > \tau_k$  are i.i.d. uniform(0, 1). By the  $L^p$ LLN we have [using the convention  $0/0 = 0$  for  $S_{n,k}/(n - \tau_k)$  when  $n = \tau_k$ ]

$$(3.10) \quad \mathbf{E} \left[ \left| \frac{S_{n,k}}{n - \tau_k} - I_k \right|^p \middle| C_k \right] \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \rightarrow \infty$$

since, with  $U$  uniformly distributed and independent of all the  $U_i$ 's,

$$(3.11) \quad \mathbf{E}[\mathbf{1}(L_{k-1} < U < R_{k-1}) \beta(U, U_{\tau_k}) \mid C_k] = I_k.$$

For our conditional application of the  $L^p$ LLN in (3.10), it is sufficient to assume only that the probabilistic source and the cost function  $\beta \geq 0$  are such that  $I_{p,k}$  is a.s. finite, and this clearly holds by (3.9).

Our next goal is to show that the left side of (3.10) is dominated by a single random variable (depending on the fixed value of  $k$ ) with finite expectation, and then we will apply the dominated convergence theorem. For every  $n$ , using the convexity of  $x^p$  for  $x > 0$  we obtain

$$\mathbf{E} \left[ \left| \frac{S_{n,k}}{n - \tau_k} - I_k \right|^p \middle| C_k \right] \leq 2^{p-1} \left( \mathbf{E} \left[ \left( \frac{S_{n,k}}{n - \tau_k} \right)^p \middle| C_k \right] + I_k^p \right).$$

We claim that each of the two terms multiplying  $2^{p-1}$  on the right here is bounded by  $I_{p,k}$ . First, using the triangle inequality for conditional  $L^p$ -norm given  $C_k$ , the fact that the random variables summed to obtain  $S_{n,k}$  are conditionally i.i.d. given  $C_k$ , and the definition (3.8) of  $I_{p,k}$ , we can bound the  $p$ th root of the first term by

$$\begin{aligned} & \left\{ \mathbf{E} \left[ \left( \frac{S_{n,k}}{n - \tau_k} \right)^p \middle| C_k \right] \right\}^{1/p} \\ & \leq \frac{1}{n - \tau_k} \sum_{i: \tau_k < i \leq n} \{ \mathbf{E} [ \mathbf{1}(L_{k-1} < U_i < R_{k-1}) \beta^p(U_i, U_{\tau_k}) | C_k ] \}^{1/p} \\ (3.12) \quad & = \{ \mathbf{E} [ \mathbf{1}(L_{k-1} < U < R_{k-1}) \beta^p(U, U_{\tau_k}) | C_k ] \}^{1/p} = I_{p,k}^{1/p} \end{aligned}$$

with  $U$  as at (3.11). For the second term we observe that  $[I_k / (R_{k-1} - L_{k-1})]^p$  is the  $p$ th power of the absolute value of a uniform average and so is bounded by the corresponding uniform average of absolute values of  $p$ th powers, namely,  $I_{p,k} / (R_{k-1} - L_{k-1})$ ; thus

$$(3.13) \quad I_k^p \leq (R_{k-1} - L_{k-1})^{p-1} I_{p,k} \leq I_{p,k}.$$

So we conclude that

$$\mathbf{E} \left[ \left| \frac{S_{n,k}}{n - \tau_k} - I_k \right|^p \middle| C_k \right] \leq 2^p I_{p,k}.$$

Thus it follows from  $\mathbf{E} I_{p,k} < \infty$  [which follows from (3.9)] and the dominated convergence theorem that

$$(3.14) \quad \mathbf{E} \left| \frac{S_{n,k}}{n - \tau_k} - I_k \right|^p \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Next, we will show from (3.14) that, for each  $k$ ,

$$(3.15) \quad \mathbf{E} \left| \frac{S_{n,k}}{n} - I_k \right|^p \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

by proving that

$$d_{n,k} \equiv d_{p,n,k} := \mathbf{E} \left| \frac{S_{n,k}}{n} - \frac{S_{n,k}}{n - \tau_k} \right|^p = \mathbf{E} \left( \frac{\tau_k}{n} \frac{S_{n,k}}{n - \tau_k} \right)^p$$

vanishes in the limit as  $n \rightarrow \infty$ . Indeed, the corresponding conditional expectation given  $C_k$  is

$$\mathbf{1}(\tau_k < n) \left( \frac{\tau_k}{n} \right)^p \mathbf{E} \left[ \left( \frac{S_{n,k}}{n - \tau_k} \right)^p \middle| C_k \right] \leq \mathbf{1}(\tau_k < n) \left( \frac{\tau_k}{n} \right)^p I_{p,k}$$

recalling the inequality (3.12). So again using  $\mathbf{E} I_{p,k} < \infty$  and applying the dominated convergence theorem we find that  $d_{n,k} \rightarrow 0$ , as desired.

Finally, we show that  $S_n^V/n$  converges to  $S$  in  $L^p$ . Since we have termwise  $L^p$ -convergence of  $S_n^V/n$  to  $S$  by (3.15), the triangle inequality for  $L^p$ -norm and the dominated convergence theorem for sums imply that  $S_n^V/n$  converges in  $L^p$  to  $S$  provided we can find a summable sequence  $b_k$  such that

$$\max \left\{ \sup_{n \geq 1} \left\| \frac{S_{n,k}}{n} \right\|_p, \|I_k\|_p \right\} \leq b_k.$$

But, for any  $n \geq 1$ , we have [by taking  $p$ th powers in (3.12), then taking expectations, then taking  $p$ th roots]

$$\left\| \frac{S_{n,k}}{n} \right\|_p \leq \left\| \frac{S_{n,k}}{n - \tau_k} \right\|_p \leq (\mathbf{E} I_{p,k})^{1/p}.$$

Further,  $\|I_k\|_p \leq (\mathbf{E} I_{p,k})^{1/p}$  follows from (3.13). Finally,  $b_k := (\mathbf{E} I_{p,k})^{1/p}$  is assumed to be summable. Thus  $S_n^V/n$  converges to  $S$  in  $L^p$ .  $\square$

**Remark 3.3.** Letting  $K_n$  denote the number of key comparisons required by  $\text{QuickVal}(n, \alpha)$ , we find from Theorem 3.1 with  $\beta \equiv 1$  that  $K_n/n$  converges in  $L^p$  ( $1 \leq p < \infty$ ) to

$$K := \sum_{k=0}^{\infty} (R_k - L_k).$$

(In Section 3.4, we will explicitly show the required condition that  $\mathbf{E} K < \infty$ ; see Remark 3.9.)

Suppose  $\alpha = 0$ ; then the number of key comparisons  $K_n$  for  $\text{QuickVal}(n, \alpha)$  is the same as for  $\text{QuickMin}$ . In this case Theorem 3.1 gives

$$(3.16) \quad \frac{K_n}{n} \xrightarrow{L^p} K = 1 + \sum_{k \geq 1} U_{\tau_k}$$

for  $1 \leq p < \infty$ . The limiting random variable  $K$  has mean 2 and the same so-called Dickman distribution as the perpetuity

$$(3.17) \quad 1 + \sum_{k \geq 1} U_1 \cdots U_k.$$

That (3.16)–(3.17) holds is well known (e.g., Mahmoud *et al.* [15], Hwang and Tsai [12]).

**3.3. Almost Sure Convergence of  $S_n^V/n$ .** Under a tameness assumption, we can also show that  $S_n^V/n$  converges to  $S$  almost surely. (Recall Definition 2.4.)

**Theorem 3.4.** *Suppose that the cost  $\beta$  is  $\epsilon$ -tamed for some  $\epsilon < 1/4$ . Then  $S_n^V/n$  defined at (3.5) converges to  $S$  almost surely.*

Before proving this theorem, we establish three lemmas bounding various quantities of interest.

**Lemma 3.5.** *For any  $p > 0$  and  $k \geq 1$ , we have*

$$\mathbf{E}(R_k - L_k)^p \leq \left( \frac{2 - 2^{-p}}{p + 1} \right)^k.$$

Here note that for all  $p > 0$  we have

$$(3.18) \quad 0 < \frac{2 - 2^{-p}}{p + 1} < 1.$$

*Proof.* Fix  $p > 0$  and  $k \geq 1$ . Since  $R_0 - L_0 = 1$ , it is sufficient to prove that

$$\mathbf{E}[(R_k - L_k)^p | L_{k-1}, R_{k-1}] \leq \frac{2 - 2^{-p}}{p + 1} (R_{k-1} - L_{k-1})^p.$$

Condition on  $(L_{k-1}, R_{k-1})$ ; then with  $U$  uniformly distributed over  $(L_{k-1}, R_{k-1})$  we have the stochastic inequality

$$R_k - L_k \leq_{\text{st}} \max\{U - L_{k-1}, R_{k-1} - U\}.$$

Thus for  $L_{k-1} \neq R_{k-1}$ , with

$$A_{k-1} := (L_{k-1} + R_{k-1})/2,$$

we have

$$\begin{aligned} & \mathbf{E}[(R_k - L_k)^p | L_{k-1}, R_{k-1}] \\ & \leq \mathbf{E}[(\max\{U - L_{k-1}, R_{k-1} - U\})^p | L_{k-1}, R_{k-1}] \\ & = (R_{k-1} - L_{k-1})^{-1} \left[ \int_{L_{k-1}}^{A_{k-1}} (R_{k-1} - u)^p du + \int_{A_{k-1}}^{R_{k-1}} (u - L_{k-1})^p du \right] \\ & = \frac{2 - 2^{-p}}{p + 1} (R_{k-1} - L_{k-1})^p, \end{aligned}$$

as desired.  $\square$

**Lemma 3.6.** *Suppose that the cost  $\beta$  is tamed with parameters  $\epsilon$  and  $c$ . Then for any interval  $(a, b) \subseteq (0, 1)$ , any  $t \in (a, b)$ , and any  $0 \leq q < 1/\epsilon$ , we have*

$$\int_a^b \beta^q(u, t) du \leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} (b - a)^{1 - q\epsilon}.$$

*Proof.* Using the tameness assumption, integration immediately gives

$$\int_a^b \beta^q(u, t) du \leq \frac{c^q}{1 - q\epsilon} [(t - a)^{1 - q\epsilon} + (b - t)^{1 - q\epsilon}].$$

The lemma now follows from the concavity of  $x^{1 - q\epsilon}$  for  $x > 0$ .  $\square$

The next lemma is a simple consequence of the preceding two.

**Lemma 3.7.** *Suppose that the cost  $\beta$  is tamed with parameters  $\epsilon < 1$  and  $c$ . Then for any  $k \geq 1$  and any  $q > 0$ , we have*

$$\mathbf{E} I_k^q \leq \left( \frac{2^\epsilon c}{1 - \epsilon} \right)^q \left( \frac{2 - 2^{-q(1 - \epsilon)}}{q(1 - \epsilon) + 1} \right)^{k-1},$$

and so  $\sum_k \mathbf{E} I_k^q < \infty$  geometrically quickly.

*Proof.* Recalling

$$I_k = \int_{L_{k-1}}^{R_{k-1}} \beta(u, U_{\tau_k}) du,$$

we find from Lemma 3.6 that

$$I_k \leq \left( \frac{2^\epsilon c}{1 - \epsilon} \right) (R_{k-1} - L_{k-1})^{1 - \epsilon}.$$

By application of Lemma 3.5 we thus obtain the desired bound on  $\mathbf{E} I_k^q$ . The series-convergence assertion follows from the observation (3.18).  $\square$

Now we prove Theorem 3.4.

*Proof of Theorem 3.4.* Clearly it suffices to show that

$$(3.19) \quad \frac{S_n^V}{n} - \frac{\tilde{S}_n}{n} \xrightarrow{\text{a.s.}} 0$$

and

$$(3.20) \quad \frac{\tilde{S}_n}{n} - S \xrightarrow{\text{a.s.}} 0,$$

where

$$\tilde{S}_n := \sum_{k \geq 1} (n - \tau_k)^+ I_k.$$

We tackle (3.20) first and then (3.19).

By the monotone convergence theorem,  $\tilde{S}_n/n \uparrow S$  almost surely. But from Lemma 3.7 (using only  $\epsilon < 1$ ) we have  $\mathbf{E} S = \sum_{k \geq 1} \mathbf{E} I_k < \infty$ , which implies that  $S < \infty$  almost surely. Hence (3.20) follows.

Our proof of (3.19) both is inspired by and follows along the same lines as the “fourth-moment proof” of the strong law of large numbers described in Ross [22, Chapter 8]; as in that proof, we prefer easy calculations involving fourth moments to more difficult ones involving tail probabilities—perhaps with the expense that the value  $1/4$  in the statement of Theorem 3.4 could be raised by more sophisticated arguments. For (3.19) it suffices to show that, for any  $\delta > 0$ ,

$$\mathbf{P} \left( \left| \frac{S_n^V}{n} - \frac{\tilde{S}_n}{n} \right| > \delta \text{ i.o.} \right) = 0,$$

for which it is sufficient by the first Borel–Cantelli lemma and Markov’s inequality to show that

$$(3.21) \quad \sum_{n \geq 1} \mathbf{E} \left( \frac{S_n^V}{n} - \frac{\tilde{S}_n}{n} \right)^4 < \infty.$$

Here, by the triangle inequality for the  $L^4$  norm,

$$(3.22) \quad \begin{aligned} \sum_{n \geq 1} \mathbf{E} \left( \frac{S_n^V}{n} - \frac{\tilde{S}_n}{n} \right)^4 &\leq \sum_{n \geq 1} \left[ \sum_{k \geq 1} \left\| \frac{S_{n,k}}{n} - \frac{(n - \tau_k)^+}{n} I_k \right\|_4 \right]^4 \\ &= \sum_{n \geq 1} \left[ \sum_{k \geq 1} \left\| \frac{(n - \tau_k)^+}{n} \left( \frac{S_{n,k}}{n - \tau_k} - I_k \right) \right\|_4 \right]^4, \end{aligned}$$

where we again use the convention  $0/0 = 0$  for  $S_{n,k}/(n - \tau_k)$  when  $n = \tau_k$ . As in the proof of Theorem 3.1, we let  $C_k$  denote the quadruple  $(L_{k-1}, R_{k-1}, \tau_k, U_{\tau_k})$ . Also we define

$$\tilde{I}_k := \mathbf{1}(L_{k-1} < U < R_{k-1}) \beta(U, U_{\tau_k}).$$

and

$$M_m(k) := \mathbf{E}[(\tilde{I}_k - I_k)^m | C_k],$$

where  $U$  is  $\text{unif}(0, 1)$  and independent of  $C_k$ . Then routine calculation (see Ross [22, Section 8.4]) shows that

$$\begin{aligned}
 \mathbf{E} \left[ \frac{(n - \tau_k)^+}{n} \left( \frac{S_{n,k}}{n - \tau_k} - I_k \right) \right]^4 &= \mathbf{E} \left[ \mathbf{E} \left[ \left\{ \frac{(n - \tau_k)^+}{n} \left( \frac{S_{n,k}}{n - \tau_k} - I_k \right) \right\}^4 \middle| C_k \right] \right] \\
 &= \mathbf{E} \left\{ \left[ \frac{(n - \tau_k)^+}{n} \right]^4 \left[ \frac{(n - \tau_k)^+ M_4(k) + 3(n - \tau_k)^+ (n - \tau_k - 1)^+ M_2^2(k)}{[(n - \tau_k)^+]^4} \right] \right\} \\
 &\leq \mathbf{E} \{ n^{-4} [n M_4(k) + 3n(n - 1) M_4(k)] \} \leq 3n^{-2} \mathbf{E} M_4(k),
 \end{aligned}
 \tag{3.23}$$

where the first inequality holds because  $M_4(k) \geq M_2^2(k)$ .

We will show that  $\mathbf{E} M_4(k)$  decays geometrically and then use that fact to prove (3.21). Since  $(a - b)^4 \leq 8(a^4 + b^4)$  for any real  $a$  and  $b$ , we have

$$M_4(k) \leq 8 \left( \mathbf{E} [\tilde{I}_k^4 | C_k] + I_k^4 \right).
 \tag{3.24}$$

First, using Lemma 3.7 we find (using only  $\epsilon < 1$ ) that  $\mathbf{E} I_k^4 < \infty$  decays geometrically:

$$\mathbf{E} I_k^4 \leq \left( \frac{2^\epsilon c}{1 - \epsilon} \right)^4 \left( \frac{2 - 2^{-4(1-\epsilon)}}{5 - 4\epsilon} \right)^{k-1}.
 \tag{3.25}$$

Now we analyze, in similar fashion,  $\mathbf{E} [\tilde{I}_k^4 | C_k]$  in (3.24). Using the assumption  $0 < \epsilon < 1/4$  and Lemma 3.6 we find

$$\mathbf{E} [\tilde{I}_k^4 | C_k] \leq \frac{2^{4\epsilon} c^4}{1 - 4\epsilon} (R_{k-1} - L_{k-1})^{1-4\epsilon}.$$

Applying Lemma 3.5 thus gives the geometric decay

$$\mathbf{E} \tilde{I}_k^4 \leq \frac{2^{4\epsilon} c^4}{1 - 4\epsilon} \left( \frac{2 - 2^{-(1-4\epsilon)}}{2 - 4\epsilon} \right)^{k-1}.
 \tag{3.26}$$

Therefore, it follows from (3.22)–(3.23) and (3.25)–(3.26) that (3.21) holds:

$$\sum_{n \geq 1} \mathbf{E} \left( \frac{S_n^V}{n} - \frac{\tilde{S}_n}{n} \right)^4 \leq 3 \left( \sum_{n \geq 1} n^{-2} \right) \left[ \sum_{k \geq 1} (\mathbf{E} M_4(k))^{1/4} \right]^4 < \infty.$$

This completes the proof of Theorem 3.4.  $\square$

**3.4. Computation of  $\mathbf{E} S$ : an integral expression.** In this section we derive the following simple double-integral expression for  $\mathbf{E} S$  in terms of the cost function  $\beta$ .

**Theorem 3.8.** *For any symmetric cost function  $\beta \geq 0$  we have*

$$\mathbf{E} S = 2 \int \int_{0 < u < t < 1} \beta(u, t) [(\alpha \vee t) - (\alpha \wedge u)]^{-1} du dt.$$

*Proof.* Recall that  $\mathbf{E} S = \sum_{k \geq 1} \mathbf{E} I_k$ , where

$$I_k = \int_{L_{k-1}}^{R_{k-1}} \beta(u, U_{\tau_k}) du.$$



Recall also that, for each  $k$ , the conditional distribution of  $U_{\tau_k}$  given  $L_{k-1}$  and  $R_{k-1}$  is uniform( $L_{k-1}, R_{k-1}$ ). Thus

$$\begin{aligned}
\mathbf{E} I_k &= \mathbf{E} \int_{L_{k-1}}^{R_{k-1}} (R_{k-1} - L_{k-1})^{-1} \int_{L_{k-1}}^{R_{k-1}} \beta(u, w) dw du \\
&= \int_{0 < w, u < 1} \beta(w, u) \mathbf{E}[(R_{k-1} - L_{k-1})^{-1} \mathbf{1}(L_{k-1} < u, w < R_{k-1})] dw du \\
&= 2 \int_{0 < w < u < 1} \beta(w, u) \\
&\quad \times \int_{0 \leq x < \alpha < y \leq 1} (y - x)^{-1} \mathbf{1}(x < w < u < y) \mathbf{P}(L_{k-1} \in dx, R_{k-1} \in dy) dw du.
\end{aligned}$$

Hence

$$\begin{aligned}
(3.27) \quad \mathbf{E} S &= 2 \int_{0 < w < u < 1} \beta(w, u) \\
&\quad \times \int_{0 \leq x < \alpha < y \leq 1} (y - x)^{-1} \mathbf{1}(x < w < u < y) \nu(dx, dy) dw du
\end{aligned}$$

where  $\nu$  is the measure

$$(3.28) \quad \nu(dx, dy) := \sum_{k \geq 0} \mathbf{P}(L_k \in dx, R_k \in dy).$$

As established in the Appendix in Proposition A.1, one has the tractable expression

$$\nu(dx, dy) = \delta_0(dx) \delta_1(dy) + (1 - x)^{-1} dx \delta_1(dy) + \delta_0(dx) y^{-1} dy + 2(y - x)^{-2} dx dy.$$

Using this last expression, routine calculation shows that, for  $0 < w < u < 1$ ,

$$(3.29) \quad \int_{0 \leq x < \alpha < y \leq 1} (y - x)^{-1} \mathbf{1}(x < w < u < y) \nu(dx, dy) = [(\alpha \vee u) - (\alpha \wedge w)]^{-1}.$$

Substitute (3.29) into (3.27) to complete the proof of the theorem.  $\square$

**Remark 3.9.** We now let  $\beta \equiv 1$  and use Theorem 3.8 to analyze the expectation of the number  $K_n$  of key comparisons required by **QuickVal**( $n, \alpha$ ). Then the expected value in Theorem 3.8 is

$$(3.30) \quad 2 \int \int_{0 < u < t < 1} [(\alpha \vee t) - (\alpha \wedge u)]^{-1} du dt = 2[1 - \alpha \ln \alpha - (1 - \alpha) \ln(1 - \alpha)] < \infty.$$

It follows by (3.30) that for  $\alpha = 0$  we have

$$\lim_{n \rightarrow \infty} \mathbf{E} K_n / n = 2,$$

which is well known since  $K_n$  in this case represents the number of key comparisons required by **QuickMin** applied to a file of  $n$  keys (e.g., Mahmoud *et al.* [15]). Thus we are now able to conclude that for any  $\alpha$  ( $0 \leq \alpha \leq 1$ ),  $\mathbf{E} K_n / n$  converges to the simple constant in (3.30). Also notice that we have verified the hypothesis of Theorem 3.1 for  $p = 1$  (see also 3.2) by (3.30), as we promised in Remark 3.3 that we would.

**3.5. Computation of  $\mathbf{E}S$ : a series expression.** We now restrict to the cost function  $\beta_{\text{sympb}}$  and use Theorem 3.8 to derive a series expression for  $\mathbf{E}S$ . In the notation of Section 2.1, we have

$$\frac{1}{2} \mathbf{E}S = \sum_{w \in \Sigma^*} \int_{\mathcal{T}_w} [(\alpha \vee t) - (\alpha \wedge u)]^{-1} du dt,$$

which is easily obtained by noting that for  $u < t$  we have

$$(3.31) \quad \beta(u, t) = \sum_{w \in \Sigma^*} \mathbf{1}(a_w < u < t < b_w).$$

Define

$$\mathcal{J}(w) := \int_{\mathcal{T}_w} [(\alpha \vee t) - (\alpha \wedge u)]^{-1} du dt.$$

Then routine calculation shows that

$$\mathcal{J}(w) = p_w L \left( \left| \frac{\alpha - \mu_w}{p_w} \right| \right).$$

Thus

$$(3.32) \quad \mathbf{E}S = \sum_{w \in \Sigma^*} p_w L \left( \left| \frac{\alpha - \mu_w}{p_w} \right| \right).$$

This last equation is in agreement with Theorem 2(i) of Vallée *et al.* [23] (see also their Figure 1). But, unlike in [23], our calculation requires *no assumptions* of tameness, nor even that  $\mathbf{E}S$  is finite.

#### 4. ANALYSIS OF QUICKQUANT

Following some preliminaries in Section 4.1, in Section 5 we show that a suitably defined  $S_n^Q/n$  converges in  $L^p$  to  $S$  for  $1 \leq p < \infty$  provided that the cost function  $\beta$  is  $\epsilon$ -tamed with  $\epsilon < 1/p$ ; hence  $S_n^Q/n$  and  $S_n^V/n$  have the same limiting distribution provided only that the cost function  $\beta$  is  $\epsilon$ -tamed for suitably small  $\epsilon$ . Granting that result for a moment, we can now relate three of the results obtained in Section 3 to previously known results reviewed in Section 2.2. From Remark 3.3 we recover the result of [10, Theorem 8] (in a cosmetically different, but equivalent, form; compare [9, Theorem 3]) for the limiting distribution of the number of key comparisons, and from Remark 3.9 we recover first-moment information for the same. Finally, recalling that  $L^1$ -convergence implies convergence of means, from (3.32) we recover at least the lead-order term in the asymptotics of [23] discussed at (2.4).

**4.1. Preliminaries.** We will closely follow the framework described in Section 3 for the analysis of **QuickVal** and construct a random variable, call it  $S_n^Q$ , that has the distribution of the total cost required by **QuickQuant** when applied to a file of  $n$  keys. Our goal is to show that, under suitable technical conditions,  $S_n^Q/n$  converges in  $L^p$  to  $S$  defined at (3.7).

Again, we define  $S_n^Q$  in terms of an infinite sequence  $(U_i)_{i \geq 1}$  of seeds that are i.i.d. uniform(0, 1). Let  $m_n$  (with  $m_n/n \rightarrow \alpha$ ) denote our target rank for **QuickQuant**. Let  $\tau_k(n)$  denote the index of the seed that corresponds to the  $k$ th pivot. As in Section 3.1 we will set the first pivot index  $\tau_1(n)$  to 1 rather than to a randomly chosen integer from  $\{1, \dots, n\}$ . For  $k \geq 1$ , we will use  $L_{k-1}(n)$  and  $R_{k-1}(n)$ , as defined below, to denote the lower and upper bounds, respectively, of seeds of words

that are eligible to be compared with the  $k$ th pivot. [Notice that  $\tau_k(n)$ ,  $L_k(n)$ , and  $R_k(n)$  are analogous to  $\tau_k$ ,  $L_k$ , and  $R_k$  defined in Section 3.1; see (3.1)–(3.3).] Hence we let  $L_0(n) := 0$  and  $R_0(n) := 1$ , and for  $k \geq 1$  we inductively define

$$\tau_k(n) := \inf\{i \leq n : L_{k-1}(n) < U_i < R_{k-1}(n)\},$$

and

$$\begin{aligned} L_k(n) &:= \mathbf{1}(\text{pivrank}_k(n) \leq m_n) U_{\tau_k(n)} + \mathbf{1}(\text{pivrank}_k(n) > m_n) L_{k-1}(n), \\ R_k(n) &:= \mathbf{1}(\text{pivrank}_k(n) \geq m_n) U_{\tau_k(n)} + \mathbf{1}(\text{pivrank}_k(n) < m_n) R_{k-1}(n) \end{aligned}$$

if  $\tau_k(n) < \infty$  but

$$(L_k(n), R_k(n)) := (L_{k-1}(n), R_{k-1}(n))$$

if  $\tau_k(n) = \infty$ . Here  $\text{pivrank}_k(n)$  denotes the rank of the  $k$ th pivot seed  $U_{\tau_k(n)}$  if  $\tau_k(n) < \infty$  and  $m_n$  otherwise. Recall that the infimum of the empty set is  $\infty$ ; hence  $\tau_k(n) = \infty$  if and only if  $L_{k-1}(n) = R_{k-1}(n)$ .

Using this notation, let

$$S_{n,k}^Q := \sum_{i: \tau_k(n) < i \leq n} \mathbf{1}(L_{k-1}(n) < U_i < R_{k-1}(n)) \beta(U_i, U_{\tau_k(n)})$$

be the total cost of all comparisons (for the first  $n$  keys) with the  $k$ th pivot key. Then

$$(4.1) \quad S_n^Q := \sum_{k \geq 1} S_{n,k}^Q$$

has the distribution of the total cost required by **QuickQuant**.

Notice that the expression (4.1) is analogous to (3.5). In fact, we will prove the  $L^p$ -convergence of  $S_n^Q/n$  to  $S$  by comparing the corresponding expressions for **QuickVal** and **QuickQuant**.

## 5. CONVERGENCE OF $S_n^Q/n$ IN $L^p$ FOR $1 \leq p < \infty$

The following is our main theorem regarding **QuickQuant**.

**Theorem 5.1.** *Let  $1 \leq p < \infty$ . Suppose that the cost function  $\beta$  is  $\epsilon$ -tamed with  $\epsilon < 1/p$ . Then  $S_n^Q/n$  converges in  $L^p$  to  $S$ .*

**Remark 5.2.** Note that as  $p$  increases, getting  $L^p$ -convergence requires the increasingly stronger condition  $\epsilon < 1/p$ . Thus we have convergence of moments of *all* orders provided the source is  $\gamma$ -tamed for *every*  $\gamma > 0$ —for example, if it is  $g$ -tamed as in Remark 2.5, as is true for memoryless and most Markov sources.

The proof of Theorem 5.1 will make use of the following analogue of Lemma 3.5, whose proof is essentially the same and therefore omitted.

**Lemma 5.3.** *For any  $p > 0$  and  $k \geq 1$  and  $n \geq 1$ , we have*

$$\mathbf{E}(R_k(n) - L_k(n))^p \leq \left( \frac{2 - 2^{-p}}{p + 1} \right)^k.$$

*Proof of Theorem 5.1.* Part of our strategy in proving this theorem is to compare **QuickQuant** with **QuickVal**. Hence we will frequently refer to the notation established in Section 3.1 for the analysis of **QuickVal**. For each  $k$ , observe that as  $n \rightarrow \infty$  we have

$$\tau_k(n) \xrightarrow{\text{a.s.}} \tau_k, \quad U_{\tau_k(n)} \xrightarrow{\text{a.s.}} U_{\tau_k}, \quad L_k(n) \xrightarrow{\text{a.s.}} L_k, \quad R_k(n) \xrightarrow{\text{a.s.}} R_k,$$

where  $\tau_k$ ,  $L_k$ , and  $R_k$ , are defined in Section 3.1 [see (3.1)–(3.3)]. (In fact, in each of these four cases of convergence, the left-hand side almost surely becomes equal to its limit for all sufficiently large  $n$ .) Thus for each  $k \geq 1$  we have

$$(5.1) \quad S_{n,k}^Q - S_{n,k} \xrightarrow{\text{a.s.}} 0,$$

where  $S_{n,k}$  is defined at (3.4); indeed, again the difference almost surely vanishes for all sufficiently large  $n$ . In proving Theorem 3.1, we showed [at (3.15)] that

$$\frac{S_{n,k}}{n} \xrightarrow{L^p} I_k,$$

where  $I_k$  is defined at (3.6), and it is somewhat easier (by means of conditional application of the strong law of large numbers, rather than the  $L^p$  law of large numbers, together with Fubini's theorem) to show that

$$(5.2) \quad \frac{S_{n,k}}{n} \xrightarrow{\text{a.s.}} I_k.$$

Combining (5.1) and (5.2), for each  $k \geq 1$  we have

$$(5.3) \quad \frac{S_{n,k}^Q}{n} \xrightarrow{\text{a.s.}} I_k.$$

What we want to show is that

$$(5.4) \quad \frac{S_n^Q}{n} = \sum_{k \geq 1} \frac{S_{n,k}^Q}{n} \xrightarrow{L^p} \sum_{k \geq 1} I_k = S.$$

Choose any sequence  $(a_k)_{k \geq 1}$  of positive numbers summing to 1, and let  $A$  be the probability measure on the positive integers with this probability mass function. Then, once again using the fact that the  $p$ th power of the absolute value of an average is bounded by the average of  $p$ th powers of absolute values,

$$\begin{aligned} \left| \frac{S_n^Q}{n} - S \right|^p &\leq \left[ \sum_{k \geq 1} \left| \frac{S_{n,k}^Q}{n} - I_k \right| \right]^p = \left[ \sum_{k \geq 1} a_k a_k^{-1} \left| \frac{S_{n,k}^Q}{n} - I_k \right| \right]^p \\ &\leq \sum_{k \geq 1} a_k a_k^{-p} \left| \frac{S_{n,k}^Q}{n} - I_k \right|^p. \end{aligned}$$

So for (5.4) it suffices to prove that, *with respect to the product probability  $\mathbf{P} \times A$* , as  $n \rightarrow \infty$  the sequence

$$a_k^{-p} \left| \frac{S_{n,k}^Q}{n} - I_k \right|^p$$

converges in  $L^1$  to 0. What we know from (5.3) is that the sequence converges almost surely with respect to  $\mathbf{P} \times A$ .

Now almost sure convergence together with boundedness in  $L^{1+\delta}$  are, for any  $\delta > 0$ , sufficient for convergence in  $L^1$  because the boundedness condition implies uniform integrability (e.g., Chung [2, Exercise 4.5.8]). Thus our proof is reduced to showing that, for some  $q > p$ , the sequence

$$\sum_{k \geq 1} a_k^{1-q} \mathbf{E} \left| \frac{S_{n,k}^Q}{n} - I_k \right|^q$$

is bounded in  $n$ , for a suitably chosen probability mass function  $(a_k)$ . Indeed, by convexity of  $q$ th power,

$$(5.5) \quad 2^{1-q} \sum_{k \geq 1} a_k^{1-q} \mathbf{E} \left| \frac{S_{n,k}^Q}{n} - I_k \right|^q \leq \sum_{k \geq 1} a_k^{1-q} \mathbf{E} \left| \frac{S_{n,k}^Q}{n} \right|^q + \sum_{k \geq 1} a_k^{1-q} \mathbf{E} I_k^q,$$

and we will show that each sum on the right-hand side of (5.5) is bounded in order to prove the theorem. The value of  $q$  that we use can be any satisfying  $\epsilon < 1/q < 1/p$ .

First we recall from Lemma 3.7 that

$$(5.6) \quad \mathbf{E} I_k^q \leq \left( \frac{2^\epsilon c}{1-\epsilon} \right)^q \left( \frac{2 - 2^{-q(1-\epsilon)}}{q(1-\epsilon) + 1} \right)^{k-1}, \quad k \geq 1,$$

with geometric decay. Thus the second sum on the right in (5.5) is finite if the cost is  $\epsilon$ -tamed with  $\epsilon < 1$  and the sequence  $(a_k)$  is suitably chosen not to decay too quickly.

Next we analyze  $\mathbf{E}|S_{n,k}^Q/n|^q$  for the first sum on the right in (5.5). Let

$$\nu_{k-1}(n) := |\{i : L_{k-1}(n) < U_i < R_{k-1}(n), \tau_k(n) < i \leq n\}|.$$

Until further notice our calculations are done only over the event  $\{\nu_{k-1}(n) > 0\}$ . Then, bounding the  $q$ th power of the absolute value of an average by the average of  $q$ th powers of absolute values,

$$(5.7) \quad \begin{aligned} \left| \frac{S_{n,k}^Q}{n} \right|^q &= \left| \frac{1}{\nu_{k-1}(n)} \sum_{i: L_{k-1}(n) < U_i < R_{k-1}(n)} \mathbf{1}(\tau_k(n) < i \leq n) \beta(U_i, U_{\tau_k(n)}) \right|^q \\ &\quad \times \left( \frac{\nu_{k-1}(n)}{n} \right)^q \\ &\leq \frac{1}{\nu_{k-1}(n)} \sum_{i: L_{k-1}(n) < U_i < R_{k-1}(n)} \mathbf{1}(\tau_k(n) < i \leq n) \beta^q(U_i, U_{\tau_k(n)}) \\ &\quad \times \left( \frac{\nu_{k-1}(n)}{n} \right)^q. \end{aligned}$$

Let  $D_k(n)$  denote the quintuple  $(L_{k-1}(n), R_{k-1}(n), \tau_k(n), U_{\tau_k(n)}, \nu_{k-1}(n))$ , and notice that, conditionally given  $D_k(n)$ , the  $\nu_{k-1}(n)$  values  $U_i$  appearing in (5.7) are i.i.d.  $\text{unif}(L_{k-1}(n), R_{k-1}(n))$ . Using (5.7), we bound the conditional expectation of  $|S_{n,k}^Q/n|^q$  given  $D_k(n)$ . We have

$$(5.8) \quad \begin{aligned} \mathbf{E} \left[ \left| \frac{S_{n,k}^Q}{n} \right|^q \middle| D_k(n) \right] &\leq [R_{k-1}(n) - L_{k-1}(n)]^{-1} \int_{L_{k-1}(n)}^{R_{k-1}(n)} \beta^q(u, U_{\tau_k(n)}) du \\ &\quad \times \left( \frac{\nu_{k-1}(n)}{n} \right)^q. \end{aligned}$$

Under  $\epsilon$ -tameness of  $\beta$  with  $\epsilon < 1/q$ , we find from Lemma 3.6 that

$$(5.9) \quad \int_{L_{k-1}(n)}^{R_{k-1}(n)} \beta^q(u, U_{\tau_k(n)}) du \leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} [R_{k-1}(n) - L_{k-1}(n)]^{1-q\epsilon}.$$

From (5.8)–(5.9), it follows that if  $\epsilon < 1/q$ , then

$$\mathbf{E} \left[ \left| \frac{S_{n,k}^Q}{n} \right|^q \middle| D_k(n) \right] \leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} [R_{k-1}(n) - L_{k-1}(n)]^{q-q\epsilon} \left( \frac{\nu_{k-1}(n)}{n(R_{k-1}(n) - L_{k-1}(n))} \right)^q.$$

Until this point we have worked only over the event  $\{\nu_{k-1}(n) > 0\}$ , but now we enlarge our scope to the event  $\{L_{k-1}(n) < R_{k-1}(n)\}$  and note that the preceding inequality holds there, as well.

Next notice that, conditionally given the triple

$$\tilde{D}_k(n) := (L_{k-1}(n), R_{k-1}(n), \tau_k(n)),$$

the values  $U_i$  with  $\tau_k(n) < i \leq n$  are i.i.d.  $\text{unif}(0, 1)$ , and so the number of them falling in the interval  $(L_{k-1}(n), R_{k-1}(n))$  is distributed binomial( $m, t$ ) with  $m = n - \tau_k(n)$  and  $t = R_{k-1}(n) - L_{k-1}(n)$ , and (representing a binomial as a sum of independent Bernoulli random variables and applying the triangle inequality for  $L^q$ ) moment of order  $q$  bounded by  $m^q t$ . Thus

$$\mathbf{E} \left[ \left( \frac{\nu_{k-1}(n)}{n(R_{k-1}(n) - L_{k-1}(n))} \right)^q \middle| \tilde{D}_k(n) \right] \leq [R_{k-1}(n) - L_{k-1}(n)]^{1-q},$$

so that

$$\mathbf{E} \left[ \left| \frac{S_{n,k}^Q}{n} \right|^q \middle| \tilde{D}_k(n) \right] \leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} [R_{k-1}(n) - L_{k-1}(n)]^{1-q\epsilon}.$$

Since this inequality holds even when  $L_{k-1}(n) = R_{k-1}(n)$ , we can take expectations to conclude

$$\begin{aligned} \mathbf{E} \left| \frac{S_{n,k}^Q}{n} \right|^q &\leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} \mathbf{E}[R_{k-1}(n) - L_{k-1}(n)]^{1-q\epsilon} \\ (5.10) \quad &\leq \frac{2^{q\epsilon} c^q}{1 - q\epsilon} \left( \frac{2 - 2^{-(1-q\epsilon)}}{2 - q\epsilon} \right)^{k-1}, \end{aligned}$$

where at the second inequality we have employed Lemma 5.3.

From (5.6) and (5.10) we see that we can choose  $(a_k)$  to be the geometric distribution  $a_k = (1 - \theta)\theta^{k-1}$ ,  $k \geq 1$ , with

$$\frac{2 - 2^{-q(1-\epsilon)}}{q(1-\epsilon) + 1} < \theta < 1.$$

We then conclude that  $\sum_{k \geq 1} a_k^{1-q} \mathbf{E} \left| (S_{n,k}^Q/n) - I_k \right|^q$  is bounded in  $n$ , and therefore that  $S_n^Q/n$  converges to  $S$  in  $L^p$ , if the cost function is  $\epsilon$ -tamed with  $\epsilon < 1/p$ .  $\square$

**Acknowledgments.** The first author carried out his research as a Ph.D. student in the Department of Applied Mathematics and Statistics at The Johns Hopkins University. We are grateful to Brigitte Vallée for suggesting Lemmas 3.6–3.7 and for a multitude of other helpful comments.

A. APPENDIX: A TRACTABLE EXPRESSION FOR THE MEASURE  $\nu$ 

The purpose of this appendix is to prove the following proposition used in the computation of  $\mathbf{E}S$  in Section 3.4.

**Proposition A.1.** *With  $(L_k, R_k)$  defined at (3.2)–(3.3) as the interval of values eligible to be compared with the  $k$ th pivot chosen by `QuickVal`, and with*

$$\nu(dx, dy) := \sum_{k \geq 0} \mathbf{P}(L_k \in dx, R_k \in dy)$$

as defined at (3.28), we have

$$\nu(dx, dy) = \delta_0(dx) \delta_1(dy) + (1-x)^{-1} dx \delta_1(dy) + \delta_0(dx) y^{-1} dy + 2(y-x)^{-2} dx dy.$$

*Proof.* To begin, since  $L_0 := 0$  and  $R_0 := 1$  we have

$$(A.1) \quad \mathbf{P}(L_0 \in dx, R_0 \in dy) = \delta_0(dx) \delta_1(dy),$$

where  $\delta_z$  denotes the probability measure concentrated at  $z$ . Now assume  $k \geq 1$ .

If  $0 \leq \lambda < \alpha < \rho \leq 1$ , then

$$\begin{aligned} \mathbf{P}(L_k \in dx, R_k \in dy \mid L_{k-1} = \lambda, R_{k-1} = \rho) \\ = \delta_\rho(dy) \mathbf{1}(\lambda < x < \alpha)(\rho - \lambda)^{-1} dx + \delta_\lambda(dx) \mathbf{1}(\alpha < y < \rho)(\rho - \lambda)^{-1} dy. \end{aligned}$$

Hence

$$(A.2) \quad \begin{aligned} \mathbf{P}(L_k \in dx, R_k \in dy) &= \int [\delta_\rho(dy) \mathbf{1}(\lambda < x < \alpha)(\rho - \lambda)^{-1} dx \\ &\quad + \delta_\lambda(dx) \mathbf{1}(\alpha < y < \rho)(\rho - \lambda)^{-1} dy] \mathbf{P}(L_{k-1} \in d\lambda, R_{k-1} \in d\rho). \end{aligned}$$

We can infer [and inductively prove using (A.2)] that, for  $k \geq 1$ ,

$$(A.3) \quad \mathbf{P}(L_k \in dx, R_k \in dy) = \delta_1(dy) f_k(x) dx + \delta_0(dx) g_k(y) dy + h_k(x, y) dx dy,$$

where

$$f_1(x) = \mathbf{1}(0 \leq x < \alpha), \quad g_1(y) = \mathbf{1}(\alpha < y \leq 1), \quad h_1(x, y) = 0,$$

and, for  $k \geq 2$ ,

$$(A.4) \quad f_k(x) = \mathbf{1}(0 \leq x < \alpha) \int \mathbf{1}(0 \leq \lambda < x)(1 - \lambda)^{-1} f_{k-1}(\lambda) d\lambda,$$

$$(A.5) \quad g_k(y) = \mathbf{1}(\alpha < y \leq 1) \int \mathbf{1}(y < \rho \leq 1) \rho^{-1} g_{k-1}(\rho) d\rho,$$

$$(A.6) \quad \begin{aligned} h_k(x, y) &= \mathbf{1}(0 \leq x < \alpha < y \leq 1) \left[ (1-x)^{-1} f_{k-1}(x) + y^{-1} g_{k-1}(y) \right. \\ &\quad + \int \mathbf{1}(0 \leq \lambda < x)(y - \lambda)^{-1} h_{k-1}(\lambda, y) d\lambda \\ &\quad \left. + \int \mathbf{1}(y < \rho \leq 1)(\rho - x)^{-1} h_{k-1}(x, \rho) d\rho \right]. \end{aligned}$$

Henceforth suppose  $0 \leq x < \alpha < y \leq 1$ . From (A.5) we obtain

$$(A.7) \quad g_k(y) = \frac{(-\ln y)^{k-1}}{(k-1)!}, \quad k \geq 1,$$

whence

$$(A.8) \quad \sum_{k \geq 1} g_k(y) = y^{-1}.$$



By recognizing symmetry between (A.4) and (A.5), we also find

$$(A.9) \quad f_k(x) = \frac{[-\ln(1-x)]^{k-1}}{(k-1)!}, \quad k \geq 1,$$

and so

$$(A.10) \quad \sum_{k \geq 1} f_k(x) = (1-x)^{-1}.$$

In order to compute  $\sum_{k \geq 1} h_k(x, y)$ , we consider the generating function

$$(A.11) \quad H(x, y, z) := \sum_{k \geq 1} h_k(x, y) z^k.$$

From (A.6),

$$(A.12) \quad \begin{aligned} H(x, y, z) = & z \left[ (1-x)^{-1} \sum_{k \geq 1} f_k(x) z^k + y^{-1} \sum_{k \geq 1} g_k(y) z^k \right. \\ & \left. + \int_0^x (y-\lambda)^{-1} H(\lambda, y, z) d\lambda + \int_y^1 (\rho-x)^{-1} H(x, \rho, z) d\rho \right]. \end{aligned}$$

Using this integral equation, we will show via a series of lemmas culminating in Lemma A.10 that

$$(A.13) \quad H(x, y) := H(x, y, 1) = \sum_{k \geq 1} h_k(x, y) \text{ equals } 2(y-x)^{-2}.$$

Combining equations (A.3), (A.8), (A.10), and (A.13), we obtain the desired expression for  $\nu$ .  $\square$

Throughout the remainder of this appendix, whenever we refer to  $H(x, y)$  we tacitly suppose that  $0 \leq x < \alpha < y \leq 1$ .

**Lemma A.2.**  $H(x, y) < \infty$  almost everywhere.

*Proof.* We revisit Remarks 3.3 and 3.9 and consider the number of key comparisons required by **QuickVal**( $n, \alpha$ ). As shown at (3.30), we have  $\mathbf{E} S < \infty$  in this case. On the other hand, with  $\beta \equiv 1$ , from (3.27)–(3.28), (A.1), (A.3), and (A.8)–(A.10), we have

$$\begin{aligned} \mathbf{E} S = & 2 \int_{0 < w < u < 1} \left[ 1 + \int_{0 \leq x < \alpha} (1-x)^{-1} \mathbf{1}(x < w) dx + \int_{\alpha < y \leq 1} y^{-1} \mathbf{1}(y > u) dy \right. \\ & \left. + \int_{0 \leq x < \alpha < y \leq 1} (y-x)^{-1} \mathbf{1}(x < w < u < y) H(x, y) dx dy \right] dw du. \end{aligned}$$

Thus  $H(x, y) < \infty$  almost everywhere.  $\square$

The next lemma establishes monotonicity properties of  $H(x, y)$ .

**Lemma A.3.**  $H(x, y)$  is increasing in  $x$  and decreasing in  $y$ .

*Proof.* For each  $k \geq 1$ , we see from (A.9) that  $f_k(x)$  is increasing in  $x$  and from (A.7) that  $g_k(y)$  is decreasing in  $y$ . Since  $h_1 \equiv 0$ , it follows by induction on  $k$  from (A.6) that  $h_k(x, y)$  is increasing in  $x$  and decreasing in  $y$  for each  $k$ . Thus  $H(x, y) = \sum_{k \geq 1} h_k(x, y)$  enjoys the same monotonicity properties.  $\square$

**Lemma A.4.**  $H(x, y) < \infty$  for all  $x$  and  $y$ .

*Proof.* This is immediate from Lemmas A.2–A.3.  $\square$

**Lemma A.5.** *The generating function  $H(x, y, z)$  at (A.11), is (with  $h_0 := 0$ ) the unique power-series solution  $\tilde{H}(x, y, z) = \sum_{k \geq 0} \tilde{h}_k(x, y) z^k$  (in  $0 \leq z \leq 1$ ) to the integral equation (A.12) such that  $0 \leq \tilde{h}_k(x, y) \leq h_k(x, y)$  for all  $k, x, y$ .*

*Proof.* We have already seen that  $H$  is such a solution. Conversely, if  $\tilde{H}$  is such a solution, then equating coefficients of  $z^k$  in the integral equation [which is valid because we know by Lemma A.4 that  $H(x, y, z)$ , and hence also  $\tilde{H}(x, y, z)$ , is finite for  $0 \leq z \leq 1$ ] we find that the functions  $\tilde{h}_k(x, y)$  satisfy  $\tilde{h}_k \equiv 0$  for  $k = 0, 1$  and the recurrence relation (A.6) for  $k \geq 2$ . It then follows by induction that  $\tilde{h}_k(x, y) = h_k(x, y)$  for all  $k, x, y$ .  $\square$

Next we let  $H_0(x, y, z) := 0$  and, for  $0 \leq z \leq 1$ , inductively define  $H_n(x, y, z)$  by applying successive substitutions to the integral equation (A.12); that is, for each  $n \geq 1$  we define

$$(A.14) \quad \begin{aligned} H_n(x, y, z) := & z \left[ (1-x)^{-1} \sum_{k \geq 1} f_k(x) z^k + y^{-1} \sum_{k \geq 1} g_k(y) z^k \right. \\ & \left. + \int_0^x (y-\lambda)^{-1} H_{n-1}(\lambda, y, z) d\lambda + \int_y^1 (\rho-x)^{-1} H_{n-1}(x, \rho, z) d\rho \right]. \end{aligned}$$

Let  $[z^k] H_n(x, y, z)$  denote the coefficient of  $z^k$  in  $H_n(x, y, z)$ .

**Lemma A.6.** *For each  $k \geq 1$ ,  $[z^k] H_n(x, y, z)$  is nondecreasing in  $n \geq 0$ .*

*Proof.* The inequality  $[z^k] H_n(x, y, z) \geq [z^k] H_{n-1}(x, y, z)$  is proved easily by induction on  $n \geq 1$ .  $\square$

According to the next lemma,  $H$  dominates each  $H_n$ .

**Lemma A.7.** *For all  $n \geq 0$  and  $k \geq 1$  we have*

$$(A.15) \quad 0 \leq [z^k] H_n(x, y, z) \leq h_k(x, y, z).$$

*Proof.* Lemma A.6 establishes the first inequality, and the second is proved easily by induction on  $n$ .  $\square$

Lemmas A.5–A.7 lead to the following lemma:

**Lemma A.8.** *For  $0 \leq x < \alpha < y \leq 1$  and  $0 \leq z \leq 1$  we have*

$$H_n(x, y, z) \uparrow H(x, y, z) \text{ as } n \uparrow \infty.$$

*Proof.* Recalling Lemmas A.6–A.7, define  $\tilde{H}(x, y, z)$  to be the power series in  $z$  with coefficient of  $z^k$  equal to  $\tilde{h}_k(x, y) := \lim_{n \uparrow \infty} [z^k] H_n(x, y, z)$ , which satisfies  $0 \leq \tilde{h}_k(x, y) \leq h_k(x, y)$ . On the other hand,  $\tilde{H}$  satisfies the integral equation (A.12) by applying the monotone convergence theorem to (A.14). Thus it follows from Lemma A.5 that  $\tilde{H} = H$ . Finally, another application of the monotone convergence theorem shows that  $\tilde{H}(x, y, z) = \lim_{n \uparrow \infty} H_n(x, y, z)$ .  $\square$

Our next lemma, when combined with the preceding one, immediately leads to inequality in one direction in (A.13).

**Lemma A.9.** For  $0 \leq x < \alpha < y \leq 1$  and all  $n \geq 0$ ,

$$H_n(x, y, 1) \leq 2(y - x)^{-2}.$$

*Proof.* We will prove this lemma by induction on  $n$ , starting with

$$H_0(x, y) = 0 \leq 2(y - x)^{-2}.$$

Suppose that the claim holds for  $n - 1$ . Then from (A.14), (A.8), and (A.10) we have

$$\begin{aligned} H_n(x, y, 1) &\leq (1 - x)^{-2} + y^{-2} + (y - x)^{-2} - y^{-2} + (y - x)^{-2} - (1 - x)^{-2} \\ &= 2(y - x)^{-2}. \end{aligned} \quad \square$$

Finally we are ready to prove (A.13).

**Lemma A.10.** For  $0 \leq x < \alpha < y \leq 1$ ,

$$H(x, y, 1) = 2(y - x)^{-2}.$$

*Proof.* Define

$$\widehat{H}(x, y) := 2(y - x)^{-2} - H(x, y).$$

Then to prove the desired equality it suffices to show that for any integer  $r \geq 0$  we have

$$(A.16) \quad 0 \leq \widehat{H}(x, y) \leq \left(\frac{2}{3}\right)^r \times 2(y - x)^{-3}.$$

As remarked earlier, the nonnegativity of  $\widehat{H}$  follows from Lemmas A.8–A.9. We prove the upper bound on  $\widehat{H}$  in (A.16) by induction on  $r$ . The bound clearly holds for  $r = 0$ . Notice that by substituting  $z = 1$  and  $H(x, y) = 2(y - x)^{-2} - \widehat{H}(x, y)$  into the integral equation (A.12) we find

$$\begin{aligned} \widehat{H}(x, y) &= 2(y - x)^{-2} - H(x, y) \\ &= 2(y - x)^{-2} - \left\{ (1 - x)^{-2} + y^{-2} + \int_0^x (y - \lambda)^{-1} [2(y - \lambda)^{-2} - \widehat{H}(\lambda, y)] d\lambda \right. \\ &\quad \left. + \int_y^1 (\rho - x)^{-1} [2(\rho - x)^{-2} - \widehat{H}(x, \rho)] d\rho \right\} \\ &= \int_0^x (y - \lambda)^{-1} \widehat{H}(\lambda, y) d\lambda + \int_y^1 (\rho - x)^{-1} \widehat{H}(x, \rho) d\rho. \end{aligned}$$

Thus if we assume that the upper bound in (A.16) holds for  $r - 1$ , then

$$\begin{aligned} \widehat{H}(x, y) &\leq \left(\frac{2}{3}\right)^{r-1} \times 2 \left[ \int_0^x (y - \lambda)^{-4} d\lambda + \int_y^1 (\rho - x)^{-4} d\rho \right] \\ &\leq \left(\frac{2}{3}\right)^r \times 2(y - x)^{-3}. \end{aligned}$$

Hence (A.16) holds for any nonnegative integer  $r$ .  $\square$

## REFERENCES

- [1] P. Billingsley. *Probability and Measure*. Wiley Interscience, New York, 3rd edition, 1995.
- [2] K. L. Chung. *A Course in Probability Theory*. Academic Press, London, 3rd edition, 2001.
- [3] J. Clément, P. Flajolet, and B. Vallée. Dynamical sources in information theory: a general analysis of trie structures. *Algorithmica*, 29(1-2):307–369, 2001. Average-case analysis of algorithms (Princeton, NJ, 1998).
- [4] L. Devroye. Exponential bounds for the running time of a selection algorithm. *Journal of Computer and System Sciences*, 29:1–7, 1984.
- [5] L. Devroye. On the probabilistic worst-case time of “Find”. *Algorithmica*, 31:291–303, 2001.
- [6] J. A. Fill. Distributional convergence for the number of symbol comparisons used by Quicksort (Extended Abstract). *Proceedings of the 21st International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (eds.: M. Drmota and B. Gittenberger)*, pages 221–234, 2010.
- [7] J. A. Fill and S. Janson. The number of bit comparisons used by Quicksort: An average-case analysis. *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 293–300, 2004.
- [8] J. A. Fill and T. Nakama. Analysis of the expected number of bit comparisons required by Quickselect. *Algorithmica*, 58:730–769, 2010.
- [9] R. Grübel. Hoare’s selection algorithm: a Markov chain approach. *Journal of Applied Probability*, 35:36–45, 1998.
- [10] R. Grübel and U. Rösler. Asymptotic distribution theory for Hoare’s selection algorithm. *Advances in Applied Probability*, 28:252–269, 1996.
- [11] C. A. R. Hoare. Find (algorithm 65). *Communications of the ACM*, 4:321–322, 1961.
- [12] H. Hwang and T. Tsai. Quickselect and the Dickman function. *Combinatorics, Probability and Computing*, 11:353–371, 2002.
- [13] D. E. Knuth. Mathematical analysis of algorithms. In *Information Processing 71 (Proceedings of IFIP Congress, Ljubljana, 1971)*, pages 19–27. North-Holland, Amsterdam, 1972.
- [14] J. Lent and H. M. Mahmoud. Average-case analysis of multiple Quickselect: An algorithm for finding order statistics. *Statistics and Probability Letters*, 28:299–310, 1996.
- [15] H. M. Mahmoud, R. Modarres, and R. T. Smythe. Analysis of Quickselect: An algorithm for order statistics. *RAIRO Informatique Théorique et Applications*, 29:255–276, 1995.
- [16] H. M. Mahmoud and R. T. Smythe. Probabilistic analysis of multiple Quickselect. *Algorithmica*, 22:569–584, 1998.
- [17] Volkert Paulsen. The moments of FIND. *J. Appl. Probab.*, 34(4):1079–1082, 1997.
- [18] H. Prodinger. Multiple Quickselect—Hoare’s Find algorithm for several elements. *Information Processing Letters*, 56:123–129, 1995.
- [19] M. Régnier. A limiting distribution of Quicksort. *RAIRO Informatique Théorique et Applications*, 23:335–343, 1989.
- [20] U. Rösler. A limit theorem for Quicksort. *RAIRO Informatique Théorique et Applications*, 25:85–100, 1991.
- [21] U. Rösler and L. Rüschendorf. The contraction method for recursive algorithms. *Algorithmica*, 29(1):3–33, 2001.
- [22] S. Ross. *A First Course in Probability*. Prentice Hall, Upper Saddle River, NJ, 6th edition, 2002.
- [23] B. Vallée, J. Clément, J. A. Fill, and P. Flajolet. The number of symbol comparisons in Quicksort and Quickselect. To appear in *36th International Colloquium on Automata, Languages and Programming (ICALP 2009)*, 2009.

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS, THE JOHNS HOPKINS UNIVERSITY,  
3400 N. CHARLES STREET, BALTIMORE, MD 21218-2682 USA

E-mail address: jimfill@jhu.edu

URL: <http://www.ams.jhu.edu/~fill/>

EUROPEAN CENTRE FOR SOFT COMPUTING, EDIFICIO DE INVESTIGACIÓN, CALLE GONZALO  
GUTIÉRREZ QUIRÓS S/N, 33600 MIERES, ASTURIAS, SPAIN

E-mail address: takehiko.nakama@softcomputing.es

URL: <http://www.softcomputing.es/>